

UNIT #12

STATISTICS

Lesson #1 – Summation Notation

Lesson #2 – Statistics Review

Lesson #3 – Standard Deviation

Lesson #4 – The Normal Distribution





SUMMATION NOTATION

ALGEBRA 2 WITH TRIGONOMETRY

Much of our work in this unit and the next will deal with situations where a large quantity of numbers will need to be added. In order to specify this addition or summarize it, we introduce a new notation, known as **summation or sigma notation** that will represent these sums.

SUMMATION (SIGMA) NOTATION

$$\sum_{i=a}^n f(i) = f(a) + f(a+1) + f(a+2) + \cdots + f(n)$$

where i is called the **index variable**, which starts at a value of a , ends at a value of n , and moves by unit increments.

Exercise #1: Evaluate each of the following sums.

(a) $\sum_{i=3}^5 2i$

(b) $\sum_{k=-1}^3 k^2$

(c) $\sum_{j=-2}^2 2^j$

(d) $\sum_{i=1}^5 (-1)^i$

(e) $\sum_{k=0}^4 (2k+1)$

(f) $\sum_{i=1}^3 i(i+1)$

Exercise #2: Which of represents the value of $\sum_{i=1}^4 \frac{1}{i}$?

(1) $\frac{1}{10}$

(3) $\frac{25}{12}$

(2) $\frac{9}{4}$

(4) $\frac{31}{24}$



Exercise #3: Consider the sum given by $\sum_{i=0}^3 {}_3C_i x^{3-i} y^i$.

(a) Write out this summation in simplest form.

(b) Write a binomial expression that this sum represents.

It is also good to be able to place sums into sigma notation. These answers, though, will not be unique.

Exercise #4: Express each sum using sigma notation. Use i as your index variable.

(a) $9 + 16 + 25 + \dots + 100$

(b) $-6 + -3 + 0 + 3 + \dots + 15$

(c) $\frac{1}{25} + \frac{1}{5} + 1 + 5 + \dots + 625$

Exercise #5: Which of the following represents the sum $3 + 6 + 12 + 24 + 48$?

(1) $\sum_{i=1}^5 3^i$

(3) $\sum_{i=0}^4 6^{i-1}$

(2) $\sum_{i=0}^4 3(2)^i$

(4) $\sum_{i=3}^{48} i$

Exercise #6: Some sums are more interesting than others. Determine the value of $\sum_{i=1}^{99} \left(\frac{1}{i} - \frac{1}{i+1} \right)$. Show your reasoning. This is known as a **telescoping series (or sum)**.



SUMMATION NOTATION
ALGEBRA 2 WITH TRIGONOMETRY - HOMEWORK

SKILLS

1. Evaluate each of the following. Place any non-integer answer in simplest rational form.

(a) $\sum_{i=2}^5 4i$

(b) $\sum_{k=0}^3 (k^2 + 1)$

(c) $\sum_{j=-2}^0 (2j + 1)$

(d) $\sum_{i=-1}^3 2^i$

(e) $\sum_{k=0}^4 (-1)^{2k+1}$

(f) $\sum_{i=1}^3 \log(10^i)$

(g) $\sum_{n=1}^4 \frac{n}{n+1}$

(h) $\sum_{r=0}^4 {}_4C_r$

(i) $\sum_{k=0}^3 256^{\frac{1}{2^k}}$

2. Which of the following is the value of $\sum_{k=0}^4 (4k + 1)$?

(1) 53

(3) 37

(2) 45

(4) 80

3. The sum $\sum_{i=4}^7 2^{i-7}$ is equal to

(1) $\frac{15}{8}$

(3) $\frac{3}{4}$

(2) $\frac{3}{2}$

(4) $\frac{7}{8}$



4. Write each of the following sums using sigma notation. Use k as your index variable. Note, there are many correct ways to write each sum (and even more incorrect ways).

(a) $-2 + 4 + -8 + \dots + 64 + -128$

(b) $\frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \dots + \frac{1}{81} + \frac{1}{100}$

(c) $4 + 9 + 14 + \dots + 44 + 49$

5. Which of the following represents the sum $2 + 5 + 10 + \dots + 82 + 101$?

(1) $\sum_{j=1}^6 (4j - 3)$

(3) $\sum_{j=1}^{10} (j^2 + 1)$

(2) $\sum_{j=3}^{103} (j - 2)$

(4) $\sum_{j=0}^{11} (4^j + 1)$

REASONING

6. A curious pattern occurs when we look at the behavior of the sum $\sum_{k=1}^n (2k - 1)$.

(a) Find the value of this sum for a variety of values of n below:

$$n = 2: \sum_{k=1}^2 (2k - 1) =$$

$$n = 4: \sum_{k=1}^4 (2k - 1) =$$

$$n = 3: \sum_{k=1}^3 (2k - 1) =$$

$$n = 5: \sum_{k=1}^5 (2k - 1) =$$

(b) What types of numbers are you summing?
What types of numbers are the sums?

(c) Find the value of n such that $\sum_{k=1}^n (2k - 1) = 196$.



Name: _____

Date: _____

STATISTICS REVIEW

ALGEBRA 2 WITH TRIGONOMETRY

Statistics is the study of sets of data and how these sets vary within the set and between sets. Doctors collect data in controlled studies to determine the effectiveness of life-saving drugs. Biologists collect data to observe trends in animal populations. Advertisers collect data to determine what television programs people watch. A **statistic** is simply a measurement that tells us something about the members of a data set.

Exercise #1: Biologists are studying the weight of brown trout in upstate New York streams. They capture and weigh 15 brown trout from the Saw Kill Creek in Red Hook, New York. Their weights, accurate to the nearest hundredth of a pound, are given below:

2.45, 1.60, 1.53, 2.89, 2.85, 1.09, 1.23, 2.39, 2.97, 2.28, 1.79, 1.48, 1.62, 1.17, 1.41

- (a) Use your calculator to find the mean weight, symbolized by \bar{x} , and the median weight. Round your mean to the nearest *hundredth*.
- (b) Use your calculator to find the first and third quartile values.
- (c) Calculate the range of this data set.
- (d) Calculate the **interquartile range** of this data set.

The mean and the median are called **measures of central tendency**. The quartiles, the range, and the interquartile range are all **measures of dispersion**, or how **spread out** a data set is. We will see another important measure of dispersion in the next lesson. It is also important to be able to work with grouped data that comes in a **frequency table**.

Exercise #2: Marketers are studying the ages of teenagers using social networking sites. They survey 100 teens for a particular site and find an age distribution given in the table below.

- (a) Find the mean age of the teens using this site.
- (b) 75% of all of the data fall at or below what age?
- (c) What is the modal age of this data set?

Age (x_i)	Frequency (f_i)
13	5
14	11
15	17
16	26
17	24
18	10
19	7



Exercise #3: A children's concert was given recently where 32 adults paid \$12.00 per ticket to attend and 68 children paid \$8.00 to attend. Which of the following was the mean ticket price paid?

- (1) \$10.00 (3) \$9.25
(2) \$7.42 (4) \$9.28

Exercise #4: Which of the following formulas, written in summation notation, would represent the mean of the data set $\{x_1, x_2, \dots, x_n\}$? Explain your choice.

- (1) $\sum_{i=1}^n x_i$ (3) $n \sum_{i=1}^n x_i$
(2) $\frac{1}{n} \sum_{i=1}^n x_i^2$ (4) $\frac{1}{n} \sum_{i=1}^n x_i$

Exercise #5: The mean and median of the following ascending data set are the same. Find the value of k . Show your work.

2, 4, 6, 7, 8, 9, 11, k

Exercise #6: Naihla wants to average 30 minutes per workout for five days this week. For the first four days she exercises for 25, 38, 40, and 29 minutes. How many minutes must she exercise on the fifth day to reach her mean of 30 minutes?

Exercise #7: Freddy has received test scores of 56%, 67%, and 50% on his first three tests this marking period. If there are only two tests remaining, what is the minimum he can score on the fourth test and still have a mathematical chance of reaching a passing average of 65%?



Name: _____

Date: _____

STATISTICS REVIEW
ALGEBRA 2 WITH TRIGONOMETRY - HOMEWORK

SKILLS

1. For the data set $\{2, 3, 3, 5, 7, 11, 12\}$, which of the following is true?

(1) median < mode < mean (3) mode < median < mean

(2) mean < median < mode (4) median < mean < mode

2. The interquartile range of the data set $\{4, 7, 10, 13, 18, 22, 30\}$ is

(1) 15 (3) 7

(2) 18 (4) 10

3. Which of the following represents the mean of the data set $\{y_1, y_2, y_3\}$?

(1) $3\sum_{i=1}^3 y_i$ (3) $\sum_{i=1}^3 y_i$

(2) $\sum_{i=1}^3 y_i^3$ (4) $\frac{1}{3}\sum_{i=1}^3 y_i$

4. The mean of five numbers in a set is 12. If the set contains a total of six numbers and the mean of all six is 14, which of the following is the sixth number?

(1) 24 (3) 32

(2) 13 (4) 18

APPLICATIONS

5. Senior citizens pay \$6.00 to attend a school play, students pay \$8.00 and adults pay \$10.00. If 26 senior citizens, 110 students, and 86 adults see a play, which of the following is closest to the average ticket cost?

(1) \$7.52 (3) \$8.54

(2) \$8.26 (4) \$9.25



6. A school is tracking its freshmen attendance for the first marking period. Shown below is a table summarizing their findings for 30 freshmen in Mr. Weiler's homeroom.

(a) Find the mean and median of this data set. Round your mean to the nearest tenth.

Days Absent (x_i)	Number of Students (f_i)
0	4
1	9
2	7
3	5
4	3
9	1
15	1

(b) What effect do the two **outliers** have on the mean compared to the median?

(c) Calculate the interquartile range for this data set. Show your calculation.

7. For the last 9 years, Dennis has earned an average salary of \$52,000. Determine what he must make in his tenth year for his ten year average to be \$52,500.

8. The mean of six numbers is 19. If five of the numbers are identical and the sixth is 12 more than the common value of these five numbers, then find the sixth number. Only an algebraic solution is acceptable.



THE STANDARD DEVIATION ALGEBRA 2 WITH TRIGONOMETRY

Measurements of central tendency, such as the mean and median, are important numbers that summarize average behavior of a data set. But, it is also important to know how much the data varies in a data set. The **interquartile range** from the last lesson was a rough measurement of **dispersion**. A more important measurement of dispersion is known as the **standard deviation** and can be thought of as the average distance the points of data in a set fall away from the set's mean.

STANDARD DEVIATION FORMULA

For the data set $\{x_1, x_2, \dots, x_n\}$ the standard deviation is given by $\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$

Exercise #1: Both of the following data sets have means of 5. Calculate the standard deviation of both data sets using the formula above and round your answers to the nearest *tenth*.

Data Set #1: $\{2, 4, 5, 9\}$

Data Set #2: $\{4, 5, 5, 6\}$

In all practical applications, we will assume the use of the graphing calculator to find the standard deviation of a data set. In *Exercise #2* we return to a data set from the last lesson.

Exercise #2: Biologists are studying the weight of brown trout in upstate New York streams. They capture and weigh 15 brown trout from the Saw Kill Creek in Red Hook, New York. Their weights, accurate to the nearest hundredth of a pound, are given below:

2.45, 1.60, 1.53, 2.89, 2.85, 1.09, 1.23, 2.39, 2.97, 2.28, 1.79, 1.48, 1.62, 1.17, 1.41

(a) Determine the mean and standard deviation of the weights of the brown trout. Round both answers to the nearest *hundredth* of a pound.

(b) Use your calculator to sort the data above in ascending order. Then, using your answers from part (a), determine the percent of brown trout weights that fall within one standard deviation of the mean.



Exercise #3: Which of the following data sets would have the smallest standard deviation? Do this without the use of your calculator.

- (1) {3, 5, 9, 13, 18} (3) {1, 3, 5, 10, 15}
(2) {-4, 0, 1, 2, 5} (4) {4, 4, 4, 5, 5}
-

Samples Versus Populations – Many times when statistical studies are done, it is not possible to get a data value for each member of the **population**. In this case a **sample** must be taken. There is a slight difference between how the population standard deviation, σ_x , is calculated and how the **sample standard deviation**, s_x , is calculated. It is important in any problem to pay close attention to whether a sample or the entire population is being discussed.

Exercise #4: A car company is trying to determine the fuel efficiency, in miles per gallon, of their latest hybrid vehicle. They drive a sample of 50 of these vehicles and determine the fuel efficiency. The data set is given in the table below.

- (a) Using your calculator, determine the mean and standard deviation of the fuel efficiency. Round your standard deviation to the nearest *hundredth*.

Fuel Efficiency (x_i)	Number of Cars (f_i)
37	2
38	6
39	18
40	15
41	8
42	1

- (b) Based on (a), how many cars in this sample had a fuel efficiency more than one standard deviation **below** the mean?

- (c) The government will only allow the company to advertise a fuel efficiency for a vehicle that is at most two standard deviation above the mean. Can the company claim this car gets 42 miles per gallon? Explain.

Exercise #5: In which of the following would it be appropriate to use the entire population instead of a sample?

- (1) determining the average weight of New Yorkers
(2) determining the average time spent on cell phones by American teenagers
(3) determining the average quiz grade for a given Algebra 2 class
(4) determining the average weight of Atlantic salmon
-



Name: _____

Date: _____

THE STANDARD DEVIATION
ALGEBRA 2 WITH TRIGONOMETRY - HOMEWORK

SKILLS

1. Which of the following data sets has the smallest standard deviation?

(1) $\{3, 4, 4, 5\}$ (3) $\{-1, 1, 3, 5\}$

(2) $\{4, 7, 10, 20\}$ (4) $\{2, 5, 9, 11\}$

2. Which of the following data sets has a standard deviation of zero?

(1) $\{-2, -1, 0, 1, 2\}$ (3) $\{7, 7, 7, 7, 7\}$

(2) $\{3, 3, 4, 5, 5\}$ (4) $\{0, 0, 1, 1, 2\}$

3. For the data set $\{3, 5, 9, 12, 15, 18, 21\}$, which of the following is true?

(1) interquartile range = $2 \cdot$ standard deviation

(2) interquartile range $> 2 \cdot$ standard deviation

(3) interquartile range $< 2 \cdot$ standard deviation

(4) interquartile range = $\frac{1}{2} \cdot$ standard deviation

APPLICATIONS

4. For each of the following statistical studies, determine whether it would be more appropriate to use a sample or the entire population.

(a) Determining the average cholesterol level of 40 year old Americans

(b) Determining the average age of math teachers at Arlington High School

(c) Determining the average weight of newborn bald-headed eagles.

5. Mr. Weiler's Advanced Calculus class took a quiz with the results shown below. Determine the mean and standard deviation of the quiz grades. Round both answers to the nearest *tenth*.

56, 68, 75, 77, 78, 82, 85, 86, 88, 90, 93, 98, 100



6. A company is trying to determine how long teenagers speak on cell phones per day. They survey 100 teenagers and find the following distribution for the amount of time they spent on a cell phone for a given day.

(a) Find the mean and standard deviation of this data set. Round both answers to the nearest *tenth* of a minute.

Number of Minutes (x_i)	Number of Teens (f_i)
0	4
15	8
30	14
45	25
60	12
75	28
90	9

(b) The company is planning on charging overage fees to any person who talks for a time greater than one standard deviation above the mean. How many of these 100 teenagers would be charged an overage fee? Show the calculations that lead to your answer.

7. A farmer would like to determine the acidity of his soil before he plants crops. He takes 12 samples of soil from a particular field and measures their pH. The results are shown below.

5.7, 6.1, 5.8, 6.7, 7.2, 6.3, 6.8, 6.5, 5.9, 6.4, 6.0, 6.9

(a) Find the mean and the standard deviation for this data set. Round both answers to the nearest *hundredth*.

(b) Determine, to the nearest percent, the portion of this data set that falls within one standard deviation of the mean.

8. On a recent SAT math test, the mean was 580 with a standard deviation of 60. If Jonathan scored a 655 on this particular exam, which of the following represents the number of standard deviations Jonathan's score was above the mean?

- (1) 1 (3) 3
 (2) 1.25 (4) 2.5

9. In a particular study, the mean heart rate for patients was 82 beats per minute with a standard deviation of 8 beats per minute. If Carlos participated in the study and his heart rate was three-quarters of a standard deviation below the mean, which of the following was his heart rate, in beats per minute?

- (1) 74 (3) 76
 (2) 88 (4) 79



Name: _____

Date: _____

THE NORMAL DISTRIBUTION ALGEBRA 2 WITH TRIGONOMETRY

Many large data sets have a distribution that can be well described with what is known as **The Normal Distribution** or the **Bell Curve**. This curve, as seen in the accompanying handout to this lesson, shows the percent of a normally distributed data set that lies between half standard deviation increments from the mean.

Exercise #1: For a population that is normally distributed, find the percentage of the population that lies

(a) within one standard deviation of the mean. (b) within two standard deviations of the mean.

(c) more than three standard deviations away from the mean. (d) between one and two standard deviations above the mean.

As can be easily seen from *Exercise #1*, the majority of any normally distributed population will lie within one standard deviation of its mean and the vast majority will lie within two standard deviations. A whole variety of problems can be solved if we know that a population is normally distributed.

Exercise #2: At Arlington High School, 424 juniors recently took the SAT exam. On the math portion of the exam, the mean score was 540 with a standard deviation of 80. If the scores on the exam were normally distributed, answer the following questions.

(a) What percent of the math scores fell between 500 and 660? (b) How many scores fell between 500 and 660? Round your answer to the nearest whole number.

(c) If Evin scored a 740 on her math exam, what percent of the students who took the exam did better than her? (d) Approximately how many students did better than Evin?



Exercise #3: The heights of 16 year old teenage boys are normally distributed with a mean of 66 inches and a standard deviation of 3. If Jabari is 72 inches tall, which of the following is closest to his height's percentile rank?

(1) 85th

(3) 98th

(2) 67th

(4) 93rd

Exercise #4: The amount of soda in a standard can is normally distributed with a mean of 12 ounces and a standard deviation of 0.6 ounces. If 250 soda cans were pulled by a company to check volume, how many would be expected to have less than 11.1 ounces in them?

(1) 17

(3) 28

(2) 23

(4) 11

Exercise #5: Biologists are studying the weights of Red King Crabs in the Alaskan waters. They sample 16 crabs and compiled their weights, in pounds, as shown below.

9.8, 10.1, 11.1, 12.4, 11.8, 13.2, 12.8, 12.5, 13.7, 11.6, 13.4, 12.3, 12.6, 14.8, 14.2 15.1

(a) Determine the mean and standard deviation for this sample of crabs. Round both statistical measures to the nearest *tenth* of a pound.

(b) Why does this sample indicate that the population would be well modeled using a normal distribution? Explain. Hint – Use your calculator to sort this data in ascending order.

(c) Assuming your mean and standard deviation from part (a) apply to a normally distributed population of crabs caught in Alaska, what percent will fall between 9.6 pounds and 15.6 pounds?

(d) If fishermen must throw back any crab caught below 10.4 pounds, approximately what percent of the crabs caught will need to be thrown back if the weights are normally distributed?



Name: _____

Date: _____

THE NORMAL DISTRIBUTION
ALGEBRA 2 WITH TRIGONOMETRY - HOMEWORK

SKILLS

1. A variable is normally distributed with a mean of 16 and a standard deviation of 6. Find the percent of the data set that:
- (a) is greater than 16 (b) falls between 10 and 22 (c) is greater than 28
- (d) is less than 1 (e) falls between 4 and 19 (f) falls between 22 and 31

APPLICATIONS

2. The weights of Siamese cats are normally distributed with a mean of 6.4 pounds and a standard deviation of 0.8 pounds. If a breeder of Siamese cats has 128 in his care, how many can he expect to have weights between 5.2 and 7.6 pounds?
- (1) 106 (3) 98
- (2) 49 (4) 111 _____
3. If one quart bottles of apple juice have weights that are normally distributed with a mean of 64 ounces and a standard deviation of 3 ounces, what percent of bottles would be expected to have less than 58 ounces?
- (1) 6.7% (3) 0.6%
- (2) 15.0% (4) 2.3% _____
4. Historically daily high temperatures in July in Red Hook, New York, are normally distributed with a mean of 84°F and a standard deviation of 4°F . How many of the 31 days of July can a person expect to have temperatures above 90°F ?
- (1) 6 (3) 9
- (2) 2 (4) 4 _____



5. The weights of four year old boys are normally distributed with a mean of 38 pounds and a standard deviation of 4 pounds. Which of the following weights could represent the 90th percentile for the weight of a four year old?
- (1) 47 pounds (3) 43 pounds
- (2) 45 pounds (4) 41 pounds
- _____
6. The lengths of songs on the radio are normally distributed with a mean length of 210 seconds. If 38.2% of all songs have lengths between 194 and 226 seconds, then the standard deviation of this distribution is
- (1) 16 seconds (3) 8 seconds
- (2) 32 seconds (4) 64 seconds
- _____
7. The heights of professional basketball players are normally distributed with a standard deviation of 5 inches. If only 2.3% of all pro basketball players have heights above 7 foot 5 inches, then which of the following is the mean height of pro basketball players?
- (1) 6 feet 5 inches (3) 6 feet 10 inches
- (2) 6 feet 2 inches (4) 6 feet 7 inches
- _____
8. On a recent statewide math test, the raw score average was 56 points with a standard deviation of 18. If the scores were normally distributed and 24,000 students took the test, answer the following questions.
- (a) What percent of students scored below a 38 on the test? (b) How many students scored less than a 38?
- (c) If the state would like to scale the test so that a 90% would correspond to a raw score that is one and a half standard deviations above the mean, what raw score is needed for a 90%?
- (d) How many of the 24,000 students receive a scaled score greater than a 90%?
- (e) The state would like no more than 550 of the 24,000 students to fail the exam. What percent of the total does the 550 represent? Round to the nearest tenth of a percent.
- (f) What should the raw passing score be set at so that no more than the 550 students fail?



THE NORMAL DISTRIBUTION

BASED ON STANDARD DEVIATION



